The application of the MLR property that is most relevant to us at the moment is

Thm Suppose that the family of dsns for $X$ specified by density or pmf $f(x|\theta)$ has MLR in a statistic $T(X)$. Then for any $\alpha \in (0,1]$ $\exists$ a test of the form

$$\phi(x) = \begin{cases} 1 & \text{if } T(x) > k \quad (<) \\ \nu & \text{if } T(x) = k \\ 0 & \text{if } T(x) < k \quad (>) \end{cases}$$

with $k \in [-\infty, \infty)$ and $\nu \in [0,1]$ that is UMP size $\alpha = E_{\theta_0}\phi(X) = \pi_\phi(\theta_0)$ for testing

$$H_0 : \theta \le \theta_0 \text{ vs } H_1 : \theta > \theta_0 \quad \leftarrow A$$

$$\left( H_0 : \theta \ge \theta_0 \text{ vs } H_1 : \theta < \theta_0 \right) \quad \leftarrow B$$

---

Again, as in the archetypal example, the idea of the proof is that the MLR property lets $T(x)$ function in the place of every likelihood ratio

$$\frac{f(x|\theta_2)}{f(x|\theta_1)} \quad \text{for } \theta_2 > \theta_1$$

and see $\phi$'s that reject for large $T(x)$ to be of N-P form

<u>Example</u> $X_1, X_2, \ldots, X_{10}$ iid Poisson $\lambda$

Consider testing $H_0: \lambda \geq 1$ vs $H_1: \lambda < 1$

Note that since for a single $X_i$

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \frac{1}{x!}\exp(x\log\lambda - \lambda)$$

where $\log\lambda$ is increasing in $\lambda$, we're in an exponential family and have MLR in $T(X) = \sum X_i$ — So the theorem says that UMP tests of our hypotheses (which are of type B) can be gotten as

$$\phi(x) = \begin{cases} 1 & \text{if } \sum x < k \\ \nu & \phantom{\text{if } \sum x} = \\ 0 & \phantom{\text{if } \sum x} > \end{cases}$$

and the size of such a test is $\pi_\phi(1)$

Using the fact that $\sum X_i \sim \text{Poisson}(10\lambda)$, if $\lambda = 1$

$$P_{\lambda=1}\left[\sum X \leq 0\right] \approx .000$$

$$P_{\lambda=1}\left[\sum X_i \leq 1\right] \approx .000$$

$$P_{\lambda=1}\left[\sum X_i \leq 2\right] \approx .003$$

$$P_{\lambda=1}\left[\sum X_i \leq 3\right] \approx .010$$

$$P_{\lambda=1}\left[\sum X_i \leq 4\right] \approx .029$$

Then if, for example, I want an $\alpha = .005$ test, I realize that $k<3$ won't work and neither will $k>3$, so I use $k=3$ —

$$P_{\lambda=1}\left[\sum X_i = 3\right] \approx .010 - .003 = .007$$

and if I take $k=3$ and $\nu = \frac{2}{7}$ I have

$$\alpha = \sup_{\lambda \geq 1} \Pi_\phi(\lambda) = \Pi_\phi(1)$$
$$= E_{\lambda=1}\phi(X)$$
$$= P_{\lambda=1}\left[\sum X_i < 3\right] + \frac{2}{7} P_\lambda\left[\sum X_i = 3\right]$$
$$\approx .003 + \frac{2}{7}(.007)$$
$$\approx .005$$

and this test is UMP size $\alpha = .005$ for these hypotheses

Jan 30

The bottom line here for UMP tests is

① Available in 1-parameter MLR families where the alternative is <u>1-sided</u>

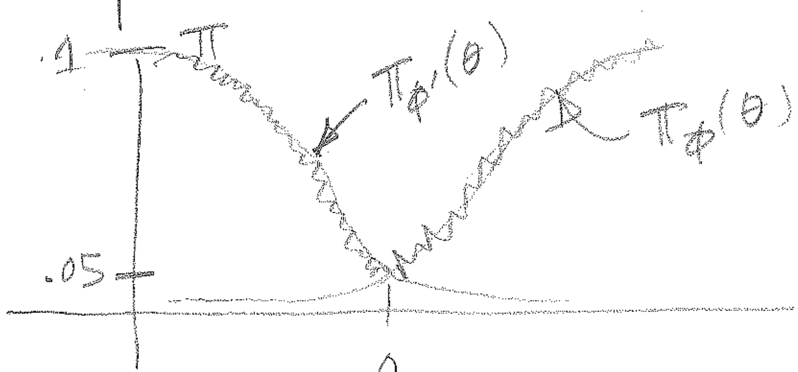
You cannot get UMP tests in MLR families for the typical Stat 101 type 2-sided alternatives

E.g. $X \sim N(\theta, 1)$ $H_0: \theta = 0$ vs $H_1: \theta \neq 0$

$\phi(x) = I[x > 1.645]$ has the largest power at any $\theta > 0$ for a test with $\pi_\phi(0) = .05$

$\phi'(x) = I[x < -1.645]$ has the largest power at any $\theta < 0$ for a test with $\pi_{\phi'}(0) = .05$

and a "uniqueness" part of N-P that I didn't state would say that these are the <u>only</u> tests that achieve their powers there




A UMP test would have to achieve the power and that's just too much to ask

(2) Available in a <u>very</u> few additional situations in which very specialized arguments can be made

E.G. In the 2-parameter problem

$$X_1, X_2, \ldots, X_n \text{ iid } N(\mu, \sigma^2)$$

with $H_0 : \sigma^2 \le \sigma_0^2$ and $H_1 : \sigma^2 > \sigma_0^2$ The usual $\chi^2$ tests are UMP

But usual $\chi^2$ tests of $H_0: \sigma^2 \geq \sigma_0^2$ vs $H_1: \sigma^2 < \sigma_0^2$

usual $t$ tests of $H_0: \mu = \mu_0$ vs $H_a: \mu \neq \mu_0$

etc

are not UMP

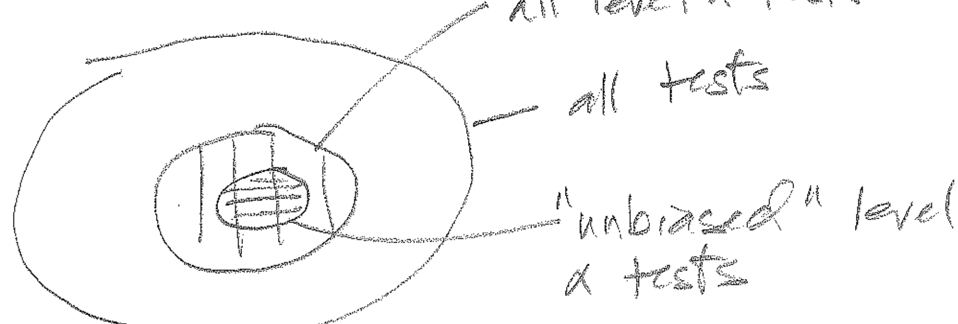
---

(3) There is a weaker and technically much harder optimality theory that can be relevant in multiparameter and 2-sided 1-parameter problems – that involves restricting the class of tests we're willing to consider – The idea is that if we're going to find a uniformly best test of $H_0: \mu = 0$ vs $H_1: \mu \neq 0$ in the $N(\mu, 1)$ model, we must disallow tests like

$$\phi(x) = I[x > 1.645]$$

$$\phi'(x) = I[x < -1.645]$$

we might do on the grounds that while their power is good on part of $\Theta_1$ it is very poor on another part of $\Theta_1$ – so the idea is




while I can't find a UMP (best) level $\alpha$ test, maybe I can find a best/UMP unbiased level $\alpha$ test

**Def** A test $\phi$ is unbiased of level $\alpha$ provided

$$\Pi_\phi(\theta) \le \alpha \text{ for } \theta \in \Theta_0$$

$$\text{and } \Pi_\phi(\theta) \ge \alpha \text{ for } \theta \in \Theta_1$$

This prevents consideration of tests whose power is awful on part of $\Theta_1$ — note that in the $N(\mu, 1)$ example with $H_0: \mu = 0$ vs $H_1: \mu \ne 0$

$$\phi(x) = I[x > 1.645]$$

is of size $\alpha = .05$, but it is not unbiased at level $\alpha$ — so it is not in the competition for the title "UMPU size $\alpha = .05$ test — This competition is won by

$$\phi'(x) = I[|x| > 1.96]$$

As it turns out, most of the usual Normal thy tests, while not UMP, do turn out to be UMPU — The theory for UMPU testing is not easy (though ultimately based on a generalization of NP) — we're going to drop the matter of optimality for testing and instead ask

What is a reasonable heuristic for test making in general contexts?

"Likelihood Ratio" Tests (This is something other than LR in NP)

Consider again $\Theta = \Theta_0 \sqcup \Theta_1$

possibly high-dimensional/complicated

## Motivations

**Simple vs Simple** N-P says optimal tests reject for large

$$R(x) = \frac{f(x|\theta_1)}{f(x|\theta_0)}$$

**Bayes** Bayes optimal tests reject for large

$$B(x) = \frac{\int_{\Theta_1} f(x|\theta) g(\theta) d\theta}{\int_{\Theta_0} f(x|\theta) g(\theta) d\theta}$$

or equivalently, reject for large

$$B'(x) = \frac{\int_{\Theta_1} f(x|\theta) \frac{g(\theta)}{P_g[\theta \in \Theta_1]} d\theta}{\int_{\Theta_0} f(x|\theta) \frac{g(\theta)}{P_g[\theta \in \Theta_0]} d\theta}$$

$$= \frac{\theta \text{ average of } f(x|\theta) \text{ according to the prior dsn conditioned to } \Theta_1}{\theta \text{ average} \qquad \Theta_0}$$

**"LRT"'s** reject $H_0$ for large

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_1} f(x|\theta)}{\sup_{\theta \in \Theta_0} f(x|\theta)}$$

Rejecting $H_0$ for $\lambda(x) > k$ is (for $k > 1$) the same as rejecting for

$$\lambda'(x) = \frac{\sup_{\theta} f(x|\theta)}{\sup_{\theta \in \Theta_0} f(x|\theta)} = \max(1, \lambda(x))$$

larger than $k$

Many common statistical procedures turn out to related to LRT's

Example $X_1, X_2, \dots, X_n$ iid $N(\mu, \sigma^2)$

$$H_0: \mu = 17 \quad \text{vs} \quad H_1: \mu \neq 17$$

$$\lambda'(x) = \frac{\sup_{\mu,\sigma^2} f(x|\mu,\sigma^2)}{\sup_{\sigma^2} f(x|17,\sigma^2)}$$

for $\quad f(x|\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp -\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$

MLEs of $\mu, \sigma^2$ are $\hat{\mu} = \bar{x}$, $\frac{n-1}{n}s^2 = \widehat{\sigma^2}$

the maximizer of $f(x|17,\sigma^2)$ is $\tilde{\sigma}^2 = \frac{1}{n}\sum(x_i-17)^2$

$$\lambda(x) = \frac{f(x|\hat{\mu},\hat{\sigma}^2)}{f(x|17,\tilde{\sigma}^2)}$$

$$= \left(\frac{\sum(x_i-17)^2}{\sum(x_i-\bar{x})^2}\right)^{n/2} \frac{\exp -\frac{1}{2\hat{\sigma}^2}\sum(x_i-\hat{\mu})^2}{\exp -\frac{1}{2\tilde{\sigma}^2}\sum(x_i-17)^2}$$

$$= \left(\frac{\sum(x_i-17)^2}{\sum(x_i-\bar{x})^2}\right)^{n/2} \frac{\exp\left(-\frac{n}{2}\right)}{\exp\left(-\frac{n}{2}\right)}$$

$$= \left(\quad\right)^{n/2}$$

Now then

$$\lambda'(x) > k \quad \text{is}$$

$$\frac{\sum (x_i - \bar{x})^2 + n(\bar{x} - 17)^2}{\sum (x_i - \bar{x})^2} > k'$$

which is

$$\sqrt{\frac{(\bar{x} - 17)^2}{\sum (x_i - \bar{x})^2}} > k''$$

which is

$$\frac{|\bar{x} - 17|}{\frac{s}{\sqrt{n}}} > k'''$$

i.e. the standard/usual $t$ test of $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ is a LRT – one chooses $k'''$ to get a desired $\alpha$ level based on the Stat 542 fact that under $H_0$

$$\frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$$

(of course standard practice in science is to try to set things up so that $\alpha$ is small)

---

The LR idea is only a heuristic and can fail to produce a sensible test ... but it "usually" leads to sensible tests in cases where I otherwise wouldn't know how to get started in a non-Bayesian way

Example $X \sim \text{Exp}$ with mean $\mu$

$H_0: \mu = 7$ vs $H_1: \mu \neq 7$

LRT rejects $H_0$ if

$$\frac{\sup_\mu \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right)}{\frac{1}{7}\exp\left(-\frac{x}{7}\right)} > k$$

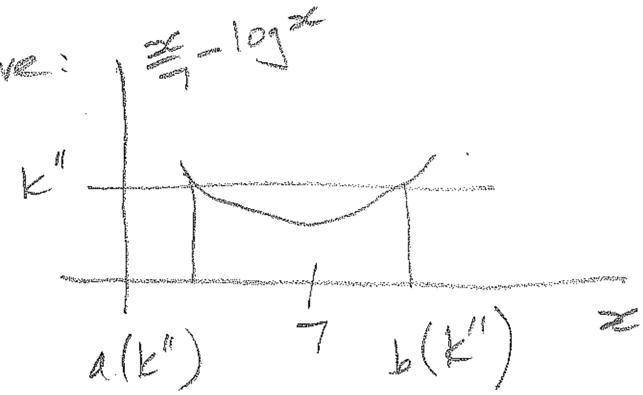
$x$ is the MLE of $\mu$ here, so this is rejection of $H_0$ if

$$\frac{7}{x} \exp\left(-1 + \frac{x}{7}\right) > k$$

i.e. if $\log 7 - \log x - 1 + \frac{x}{7} > k'$

i.e. if $\frac{x}{7} - \log x > k''$

Picture:




For a given $k''$ it is a numerical problem to find $a(k'') < 7$ and $b(k'') > 7$ solving

$$\frac{x}{7} - \log x = k''$$

Then the LRT (for $k''$) is

$$\phi(x) = I\left[x < a(k'') \text{ or } x > b(k'')\right]$$

and the size of the test is

$$P_{\mu=7}\left[\phi(X)=1\right] = 1-\left(\exp\left(-\frac{a(k'')}{7}\right)-\exp\left(-\frac{b(k'')}{7}\right)\right)$$

and by varying $k''$ we come up with a test of any desired size $\alpha$